Impact of Local Lymph Node Assay Variability on Predictions of a Bayesian Network Integrated Testing Strategy for Skin Sensitization Potency

J Pirone[1], M Smith[1], J Strickland[2], W Casey[3], J Jaworska[4]

[1]SSS, Inc., Durham, NC, USA; [2]ILS/NICEATM, RTP, NC, USA; [3]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA, [4]P&G NV, Strombeek-Bever, Belgium

Abstract

As toxicity testing moves away from traditional animal models towards cell-based assays and *in silico* methods, computational models integrating such data are being developed and improved. An example is the Bayesian network (BN) model used to predict local lymph node assay (LLNA) potency classification of substances in the NICEATM LLNA database. Datasets used to build such models may include multiple values for some combinations of assays and compounds. Using standard Bayesian network methods, it is difficult to build a model that makes use of all the available data. Instead, the data are either collapsed or selected from to produce a single value, which eliminates all distributional information. Using a published BN integrated testing strategy (ITS-2) for skin sensitization, we developed a method that incorporates the variability due to multiple LLNA values. Markov chain Monte Carlo is used to calculate results for a large number of BNs generated under distributional assumptions on the LLNA variable. This method propagates the variability through all model building steps. The most probable class predictions between the original ITS-2 and the MCMC model are similar, but the distributions of the predictions differed. These more transparent methods enhance risk assessment by describing the variability from the data and the model and better represent the reliability of the predictions.

Introduction

- It is unlikely that a single non-animal assay or *in silico* model will provide sufficient information on the risk or hazard posed by a chemical. Data from multiple inputs will need to be integrated in a way that maximizes the utility of the available information.
- Computational methods play an important role in data integration. Supervised machine learning algorithms, such as Bayesian networks, random forests, and support vector machines, which find patterns in a training dataset and use these patterns to make predictions on a new test dataset, have been widely applied (Kavlock et al. 2012).
- The training datasets used to develop these models have a single value for every predictor (e.g., assay) for each case (e.g., chemical). For some cases, only a single experimental value may be available. For others, several experimental values are reduced to a single value by data reduction (e.g., averaging) or selection.
- Collapsing data in this way completely eliminates all information on variability and may result in overly optimistic or in some cases, biased, models.
- This effect may be particularly strong when *in vitro* assay data are used to predict toxicological endpoints derived from animal models, as has been done in most published studies.
- We have previously presented a Bayesian network integrated testing strategy (ITS) for skin sensitization that avoids animal testing by using *in vitro* assays and *in silico* models (Pirone et al. 2014). Here, we build upon this model and develop new computational tools that can better incorporate the variability in the training set used to build the model.

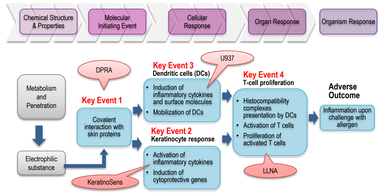
Acknowledgements

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by ILS under NIEHS contract HHSN273201400003C and SSS, Inc., under NIEHS contract GS-23F-9806H.

The views expressed above do not necessarily represent the official positions of any Federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.

A summary of NICEATM activities at SOT 2015 is available on the National Toxicology Program website at http://ntp.niehs.nih.gov/go/742110.

Figure 1. The AOP for Skin Sensitization[a]



Abbreviations: AOP = adverse outcome pathway; DPRA = direct peptide reactivity assay; LLNA = murine local lymph node assay.

[a] Adapted from OECD (2012).

Bayesian Network Integrated Testing Strategy for Skin Sensitization

- A Bayesian network is a type of probabilistic graphical model (Koller and Friedman 2009) that represents the conditional dependencies of a group of variables (e.g., assays) using a directed acyclic graph.
- The structure of the Bayesian network was designed to be consistent with the adverse outcome pathway (AOP) for substances that initiate the skin sensitization process by covalently binding to skin proteins (Jaworska et al. 2011; Jaworska et al. 2013). There are four key events in the AOP (**Figure 1**). In order of occurrence they are: 1) covalent binding to skin proteins, 2) inflammatory responses in the keratinocyte, 3) activation of dendritic cells, and 4) T-cell proliferation (OECD 2012)
  - **Table 2** links these events to the nodes (variables) found in the ITS structure (Jaworska et al. 2013) shown in **Figure 2**.
- The Bayesian network used for the skin sensitization model is discrete.
  - Associated with each node is a conditional probability table (CPT) that gives the probability of the node being in a particular state, given the values of the parent nodes.
  - For example, for the CD86 node, the associated CPT gives the probability that CD86 has a particular value given the values of the LLNA and Cysteine nodes.
- A categorical representation of a compound's potency in the murine local lymph node assay (LLNA) is used as the target endpoint. The effective concentration that produces a stimulation index of 3 (EC3), the threshold for a positive response in the LLNA, is used to describe potency. The EC3 cutoffs for the four LLNA potency categories used in the ITS are shown in **Table 1**.
- The **logKow**, **AUC120**, and **Cfree** variables are clustered to form the **Bioavailability** latent variable (**Figure 2**). Similarly, the **CD86**, **KEC3**, **KEC1.5**, **DPRACys**, and **TIMES** results are clustered to form the **Cysteine** latent variable (**Figure 2**). Latent variables increase the interpretability of the network, while at the same time reducing its computational complexity.

Table 1. LLNA EC3 Correspondence to Skin Sensitization Potency Categories

Category Number	Category Description	EC3 Range
1	Nonsensitizer	No EC3
2	Weak	EC3 ≥ 10%
3	Moderate	1% ≤ EC3 < 10%
4	Strong and extreme	EC3 < 1%

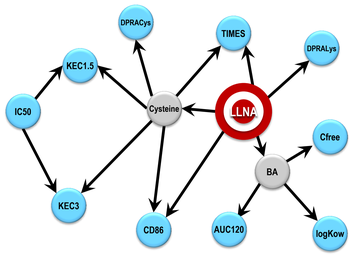
Abbreviations: EC3 = effective concentration that produces a stimulation index of 3, the threshold for a positive response in the LLNA; LLNA = murine local lymph node assay.

Table 2. Variables for the Bayesian Network ITS-2 Skin Sensitization Model

Measurement	Description	Model Variable	AOP Key Event
Physicochemical Property	Octanol–water partition coefficient	**logKow**: Log K_{ow}	Substance must penetrate the stratum corneum (step 1 of the AOP; not a key event)
Epidermal Bioavailability	Concentration of chemical reaching the mid-epidermal layer of skin calculated using a transdermal transport model (Kasting et al. 2008).	1) **Cfree:** free test substance concentration in mid-epidermis multiplied by thickness of viable epidermis (0.01 cm) expressed as percent of applied dose 2) **AUC120:** area under the flux curve at 120 h as percent of applied dose	Substance must penetrate the stratum corneum (step 1 of the AOP; not a key event)
Direct Peptide Reactivity Assay (DPRA)	*In chemico* method that measures peptide remaining after the test substance binds to two model heptapeptides	1) **DPRACys:** percent cysteine peptide remaining 2) **DPRALys:** percent lysine peptide remaining	1) Binding to skin proteins
KeratinoSens Assay	*In vitro* test that detects electrophiles using the Nrf2 electrophile-sensing pathway in the HaCaT (immortalized keratinocyte) cell line	1) **KEC1.5:** average concentration that produces 1.5-fold enhanced activity (μM) 2) **KEC3:** average concentration yielding 3-fold enhanced activity (μM) 3) **IC50:** concentration producing 50% cytotoxicity (μM)	2) Keratinocyte inflammatory responses
U937 Activation Test	*In vitro* test that uses the human myeloid cell line U937	**CD86:** EC150 (μM) for CD86 cell surface marker expression	3) Dendritic cell activation
LLNA	*In vivo* test for skin sensitization. EC3 is used to categorize potency as noted in **Table 1.**	**LLNA:** categorical representation of LLNA potency 1 = nonsensitizer 2 = weak sensitizer 3 = moderate sensitizer 4 = strong and extreme sensitizers	4) T-cell proliferation
TIMES-M	*In silico* categorical prediction of skin sensitization potency using TIMES (Tissue Metabolism Simulator) software (V.2.25.7), an expert system that makes predictions based on knowledge about the parent compound and potential skin metabolites (Dimitrov et al. 2005).	**TIMES:** three categories: nonsensitizer, weak sensitizer, and moderate/strong/extreme sensitizer	Not a key event of the AOP. Model links parent and metabolite structures to skin sensitization outcomes in animals and humans

Abbreviations: AOP = adverse outcome pathway (OECD 2012); EC150 = effective concentration that produces a 1.5-fold increase in the CD86 cell surface marker expression, the threshold for a positive response; EC3 = effective concentration that produces a stimulation index of 3, the threshold for a positive response in the LLNA; ITS-2: integrated testing strategy 2 (Jaworska et al. 2013); LLNA = murine local lymph node assay.

Figure 2. Structure of the Bayesian Network ITS-2 Skin Sensitization Model



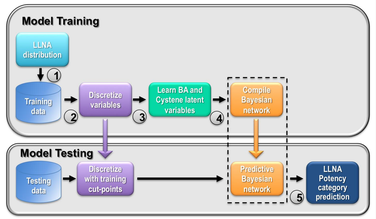
Abbreviations: BA = bioavailability; ITS-2 = integrated testing strategy 2 (Jaworska et al. 2013); abbreviations for other variables are provided in **Table 2**.

The arrows show the conditional dependencies of the variables that impact LLNA potency. BA and Cysteine are latent variables for bioavailability and cysteine binding, respectively.

Including Variability in Measured LLNA Values

- The training and test datasets from Pirone et al. (2014) were used in this analysis.
  - There are 124 chemicals in the training dataset: 36 nonsensitizers, 28 weak sensitizers, 35 moderate sensitizers, and 25 strong or extreme sensitizers.
  - There are 21 chemicals in the test dataset: 6 nonsensitizers, 5 weak sensitizers, 5 moderate sensitizers, and 5 strong or extreme sensitizers.
- Using the LLNA database compiled by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), we determined that multiple EC3 values are available for 38 chemicals in the training dataset.
  - For some chemicals, EC3 values span three potency categories. This major source of variability is not accounted for in current models (Pirone et al. 2014; Jaworska et al. 2013).
- Implementing the method to incorporate the variability due to multiple LLNA values involves several computational steps as outlined in **Figure 3**.
  - First, LLNA variability is modeled by drawing a set of plausible LLNA values for each chemical in the training dataset. A categorical-Dirichlet distribution is used; it is parameterized using the LLNA values derived from the EC3 values in the NICEATM LLNA database. For these distributions, the probability mass is greatest where there are multiple LLNA values for a chemical and is non-zero at all LLNA category levels.
  - Second, the Class-attribute Interdependence Maximization (CAIM) (Kurgan and Cios 2004) supervised discretization algorithm is used to find cut-points that bin the continuous assay data in the training data into intervals. The test data set cannot be used to find the discretization cut points, since doing so would result in biased and overly optimistic prediction results. The cut-points found for the training data are used to discretize the test data.
  - Third, mechanistically related assays are clustered to form latent (unobserved) variables.
  - Fourth, the relationships among variables in the discretized training data (including the latent variables) are described and quantitated by the dependencies implied by the Bayesian network (**Figure 2**). Each node (assay) is described by a categorical likelihood with a flat Dirichlet prior.
  - Finally, Markov chain Monte Carlo (MCMC) is used to sample from the distribution over all variables given the available evidence (posterior distribution). Predictions of LLNA potency category are made for the test set chemicals.
  - Each step in this process is repeated 1000 times giving a set of LLNA potency class predictions that reflects the variability due to the multiple LLNA potency class measurements available for the training set chemicals.
  - All computations were carried out using R version 3.1.2 (R Core Team 2013) and Jags version 3.4.0 (Plummer 2003).

Figure 3. Primary Computational Steps of the MCMC Bayesian Network Modeling Process



Abbreviations: BA = bioavailability; LLNA = murine local lymph node assay; MCMC = Markov chain Monte Carlo.

Table 3. Confusion Matrices for the Test Set with (MCMC) and without (Bayesian Network) LLNA Variability[a]

	MCMC Bayesian Network				Bayesian Network			
Experimental Value	**Nonsensitizer**	**Weak Sensitizer**	**Moderate Sensitizer**	**Strong/Extreme Sensitizer**	**Nonsensitizer**	**Weak Sensitizer**	**Moderate Sensitizer**	**Strong/Extreme Sensitizer**
Nonsensitizer	**6**	0	0	0	**6**	0	0	0
Weak Sensitizer	1	**4**	0	0	1	**4**	0	0
Moderate Sensitizer	0	0	**5**	0	0	1	**4**	0
Strong/Extreme Sensitizer	0	0	3	**2**	0	0	3	**2**

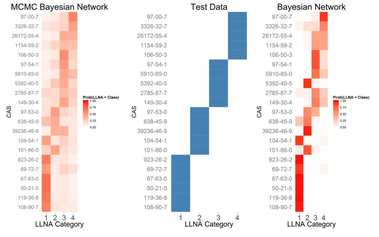
Abbreviations: LLNA = murine local lymph node assay; MCMC = Markov chain Monte Carlo.

[a] The numbers show the number of chemicals predicted in each category. Bolded numbers on the diagonals show the correct predictions.

Results

- In **Figure 4**, predictions on the test dataset obtained from the MCMC Bayesian network predictions are compared to those from the previously published Bayesian network ITS for skin sensitization (Pirone et al. 2014; Jaworska et al. 2013).
  - For each chemical in the test dataset, the probability that the chemical belonged to each of the four LLNA potency categories was determined.
  - The probabilities from the MCMC Bayesian network were obtained by pooling the results from all of the simulations.
  - For comparison, the experimentally determined LLNA potency categories are also shown. Only single values are available for chemicals in the test dataset.
  - The potency category distributions are similar for both analyses. The distributions obtained using the MCMC are less peaked, reflecting the increased variability due to the multiple LLNA potency category values for some chemicals in the training dataset.
- **Table 3** compares the most likely LLNA potency category predictions for each method to the experimentally determined value. The overall predictive accuracy is slightly better for the MCMC method, with one chemical (CASRN 5392−40−5) that was incorrectly predicted as being in potency Category 2 by the previously published approach being correctly predicted to be in potency Category 3 by the MCMC method.

Figure 4. Prediction of LLNA Potency Category for Test Set Substances[a]



Abbreviations: CAS = Chemical Abstracts Service Registry Number; LLNA = murine local lymph node assay; MCMC = Markov chain Monte Carlo.

[a] MCMC analysis includes variability in the LLNA potency measurements, while the Bayesian network analysis does not. Test Data shows the experimental results. LLNA categories are described in **Table 1**.

Conclusions

- We have presented a flexible, extensible, and fast method for incorporating variability that utilizes widely available computational tools. The MCMC ITS model predicts LLNA potency category results with 81% accuracy.
- The model better incorporates major sources of variability in the data and will result in more accurate and robust predictions.
  - Variability in experimental LLNA potency is larger for intermediate potency categories (Hoffman 2014).
  - Incorporating variability in the ITS improved the overall prediction of LLNA potency category.
    - A moderate sensitizer that was incorrectly categorized as a weak sensitizer by the previous ITS model was correctly categorized by the MCMC model.
- Current work includes:
  - Collection and curation of additional *in vivo* and *in vitro* data
  - Exploring computational methods to further decrease the time required for simulation

References

Dimitrov SD, Low LK, Patlewicz GY, Kern PS, Dimitrova GD, Comber MHI, et al. 2005. Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. Int J Toxicol 24: 189–204.

Hoffman S. LLNA variability: an essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies [Abstract]. 9th World Congress on Alternatives and Animal Use in the Life Sciences; Prague, Czech Republic, 24–28 August, 2014.

Jaworska J, Harol A, Kern PS, Gerberick GF. 2011. Integrating non-animal test information into an adaptive testing strategy—skin sensitization proof of concept case. ALTEX 28: 211–225.

Jaworska J, Dancik Y, Kern P, Gerberick GF, Natsch A. 2013. Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. J Appl Toxicol 33: 1353-1364.

Kasting GB, Miller MA, Nitsche JM. 2008. Absorption and evaporation of volatile compounds applied to skin. In: Dermatologic, Cosmeceutic and Cosmetic Development (Walters KA and Roberts MS, eds). New York: Informa Healthcare.

Kavlock R, Chandler K, Houck K, Hunter S, Judson R, Kleinstreuer N, et al. 2012. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 25: 1287–1302.

Koller D, Friedman N. 2009. Probabilistic Graphical Models. MIT Press.

Kurgan LA and Cios KJ. 2004. CAIM discretization algorithm. IEEE Transactions on Knowledge and Data Engineering, 16: 145–153.

OECD. 2012. OECD Series on Testing and Assessment No. 168. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Assessment. Paris:OECD Publishing. Available: http://www.oecd.org/env/ehs/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm [accessed 2 Dec 2013]

Pirone JR, Smith M, Kleinstreuer NC, Burns TA, Strickland J, Dancik Y, et al. 2014. Open source software implementation of an integrated testing strategy for skin sensitization potency based on a Bayesian network. ALTEX 31: 336-340.

Plummer M. 2003 Proceedings of the 3rd International Workshop on Distributed Statistical Computing.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. www.R-project.org.